University of Glasgow | College of Medical, Veterinary & Life Sciences

META-ANALYSIS WORKFLOW USED TO UNRAVEL THE MICROBIAL COMMUNITIES IN PLASTISPHERE.

A thesis submitted in partial fulfilment of the requirements for the degree of
MASTER OF SCIENCE IN BIOINFORMATICS

SUPERVISOR: Dr. Umer Zeeshan Ijaz
Co- Supervisor: Dr. Ciara Keating

SUBMITTED BY:
SHUBHRA PRIYADARSHINI
2592097P
25th October 2021

ACKNOWLEDGEMENT:

I would like to express my sincere gratitude to several individuals and organizations for supporting me throughout my post-Graduate study. Firstly, I wish to express my sincere gratitude to my Supervisors, **Dr Umer Zeeshan Ijaz** and **Dr Ciara Keating**, for their enthusiasm, patience, insightful comments, helpful information, practical advice, and unceasing ideas that have helped me tremendously at all times in my research and writing of this thesis. Their immense knowledge, profound experience and professional expertise in this topic has enabled me to complete this research successfully. Without their support and guidance, this project would not have been possible. Secondly, I would like to thank University of Glasgow for giving me an opportunity to be a part of this curriculum and helping me learn so much.

ABSTRACT:

Plastics are now undeniably pervasive and hazardous in ecosystems all around the world. Many theories have been proposed about the role of biofilms colonising plastics in the environment, called the "Plastisphere", in the transportation and environmental impact of these materials. The plastisphere, a microplastic dispersed with a growing microbial population, has a profound influence on the entire aquatic system, and its recent abundance has sparked environmental concern. We completed the meta-analysis of the Plastisphere in marine, freshwater, other aquatic habitats by collecting and re-analyzing all raw 16S rRNA gene sequencing in V4 region and metadata was collected from 842 samples across 7 published research studies. This project uses publically available data, pre-processing the data with DADA2, and then performing analysis on the data with R studio. The impact of microplastics on the marine environment and biological environment has been discovered to be significant in previous research studies.

TABLE OF CONTENTS:

1. INTRODUCTION:

1.1 BACKGROUND:

Plastic pollution is a threat to ecosystems all over the globe, with clear implications for animals like birds, marine mammals, and fish (Thiel et al.,2018). Additionally, plastic pollution potentially causes food safety problems (Smith et al., 2018), leads to developmental, reproductive, and metabolic disorders in invertebrates through the uptake of nano-sized particles (Shen et al.,2019), and even contributes to the spread of pathogens and antibiotic resistance by serving as a raft for microorganisms (Keswani et al.,2016). Plastic pollution has become a frequent environmental pollutant due to its persistence in the ecosystem; approximately 4.8-12.7 million tonnes of plastic garbage have reached the sea up to now, and plastic pollution is widely recognised as an expanding ecological threat. It is estimated that the amount of plastic released into terrestrial environments is 4–23 times higher than into marine environments (Horton et al., 2017). Plastics can reach habitats not only by direct littering and transport through rivers and oceans, but also by transport through the atmosphere and deposition in pristine regions such as alpine and Arctic environments (Allen et al., 2019). Plastic pollution has even been reported in remote areas of the Swiss Alps and the Arctic (Ambrosini et al.,2019). Developing biodegradable plastics could be one strategy to combat plastic pollution in the environment. However, they only account for about 1% of total plastic production at the moment. Furthermore, because novel materials are often tested for breakdown at temperatures above 20°C, little is known about the fate of biodegradable plastics in terms of microbial colonisation and degradability, particularly in soils from cold settings. In one study on the microbial degradation of plastic in temperate, marine sediment, there were no signs of biodegradation for both polyethylene (PE) and biodegradable carrier bags after a 3-month incubation at 10°C (Nauendorf et al.,2016).

1.2 MICROPLASTICS:

Plastics can broadly fall into four levels based on their sizes: megaplastics (> 1 m); macroplastics (< 1 m), mescoplastics (< 2.5 cm); and microplastics (< 5mm) (Gregory et al.,1996). Microplastics are a contaminant that is becoming increasingly eco-toxicologically problematic in aquatic habitats, as well as for human health. Microplastics (diameter <5 mm), which are diverse in shape, size, and density and could escape from the wastewater treatment process, result in considerable variability in distribution under spatial and temporal conditions. Micro-plastics were initially discovered in the marine environment in 1972, when a huge number of small plastic particles were discovered on the surface of the Sargasso Sea. According to Goldberg's figures, the entire amount of marine plastic garbage created around the world in 1975 was only approximately 6.4 million tonnes. Microplastics have the greatest evident influence on biota. Animals are commonly killed via physical ingestion of the digestive system and tangling of the body, and hazardous substances are also transported indirectly by modifying other species' feeding patterns. Plastic trash is non-biodegradable and lasts a long time in the natural environment because of its chemical persistence and hydrophobic character, which is dominated by polyethylene, polypropylene (floating form), PET, PVC, and other materials (sediment form).

Microplastics can take on a variety of morphologies in environmental samples, including pellets, fibres, and pieces, depending on their source.

Primary microplastics are most likely entering the aquatic environment through household sewage discharge of plastic resin powders or pellets used for air-blasting (Gregory et al.,1996). Another significant source of primary microplastics is the application of sewage sludge containing synthetic fibers or sedimented microplastics from personal care or household products to land (Horton et al.,2017).

Secondary sources of microplastics are considered as a great contributor of microplastic pollution given the large amount of macroplastic wastes entering the

environment (Duis and coors et al.,2016). Secondary microplastics, originate from anthropogenic activities, such as littering and are released during municipal solid waste collection and disposal processes (Horton et al., 2017). Wind dispersal, soil erosion, and surface runoff all have the potential to bring big plastic items and their degradation products into aquatic habitats. Similarly, wind can move light macro- and microplastics across the ground, although denser polymers are more likely to be buried deeper in soil strata (Horton et al., 2017).

Microplastics are found in many products, such as toothpaste and face scrub cream. Every year, over 280 million tonnes of plastics are discarded as waste. The accumulation of microplastics in the natural environment of fresh water, ocean, and land has grown increasingly significant. Most plastic wastes will breakdown in the presence of light, oxygen, and creatures, or will be influenced by physical forces such as waves, resulting in the formation of shattered little objects.

1.3 BACKGROUND OF PLASTIPHERE:

The word "Plastisphere" refers to the thin coating of varied metazoan and microbial life that grows on any piece of plastic in aquatic settings, just like the term "biosphere" refers to the thin layer of life on the exterior of our planet Earth. The microbial communities colonizing plastics—commonly termed the “Plastisphere” however, have only been specifically investigated more recently (Zettler et al., 2013). A call for research into the interactions between microorganisms and plastics at the beginning of 2011 marks almost a decade of Plastisphere research (Harrison JP et al., 2011). The plastic that does not reach a recycling facility or landfill, will most likely end up in our oceans due to accidental dumping of the waste, losses during transport, or direct disposal from boats (Paco, Ana et al.,2019) Plastic pollution acts as a more durable "ship" than biodegradable material for carrying the organisms over long distances (Thomas Russell et al.,2021). Because of the slow rate of biodegradation and the

varied circumstances, the plastisphere ecology varies from that formed by other floating materials that naturally exist. Insects, in addition to microorganisms, have flourished in formerly inhospitable parts of the ocean. Since the discovery of the plastisphere there has a been a multitude of research published on the topic, and many have proposed that the microbial diversity within the plastisphere is very high (Quero et al.,2017).

1.4 AMPLICON SEQUENCING OF THE16s rRNA GENE TO UNDERSTAND THE PLASTISPHERE MICROBIOTA.

At the turn of the twenty-first century, Next generation sequencing for the 16S ribosomal RNA (rRNA) research method was established, resulting in significant changes in the plastisphere of microbial ecology research. The RNA component of the 30S subunit of the bacterial ribosome is encoded by the 16S ribosomal RNA gene. Due to the difficulty of DNA–DNA hybridization, 16S rRNA gene sequencing is used to identify bacteria at the species level and to distinguish between closely related bacterial species (Wright et al.,2020). The key benefit of this methodology was that it allowed for a large increase in the creation of genuine data while lowering the costs of sample collecting. Multiplexing sequencing, immobilisation of a large library of DNA templates, and in vitro amplification were among the optimization strategies that helped to increase the quality of microbial materials analysis. In addition, Illumina sequencing approach was chosen as the best option due to its low error rate, low prices, and high flexibility in data collecting. Illumina MiSeq are mostly applied for clonal amplifications, base calling application of data analysis, variant calling, genomic DNA sequencing. The recording of DNA strand synthesis in a cluster of sample templates is part of the Illumina MiSeq sequencer method's sequencing technique. The recording technique causes the formation of freshly connected bases, which liberates the fluorescent dye that can be stimulated by diode lasers. For the purposes of presenting the process to the researchers, two digital cameras

are used. The Illumina MiSeq sequencer has numerous advantages, including variable reading length adjustment due to effective sequential interrogation of bases during a single run, a large number of samples (up to 96 samples) that may be sequenced in a single run, and high single lane flow capabilities.

The study's approach involves the use of a variety of data collection and analytic techniques. The following instrument was used for data collection: next-generation sequencing (NGS) for 16S ribosomal RNA (rRNA). All investigations focused on the V4 region because of the substantial nucleotide heterogeneity that allows for great discriminating power. Amplicon Sequence Variants (ASVs) method was utilized, which is able to gain finer resolution by observing all biological variation based on potential noise models (Callahan et al.,2017).

1.5 DIFFERENT APPROACHES OF META-ANALYSIS:

Multiple research cannot be compared at the same time because various studies used different methodologies and analyses to characterise microbial communities in different environments with varied parameters (pH, salinity, colonising plastic type, etc.). In a meta-analysis for annotated systematic review, however, we can run more statistics with a high number of samples. Secondary data on the number of studies was gathered in this way from previous research; it contributed to a complex picture of the research issue and various perspectives on its resolution. The major goal of the meta-data analysis method was to combine primary data from many studies into a unified study framework, allowing for more efficient use of existing data and expertise in the creation of trustworthy results. Understanding the microbial community structure linked with microplastics in diverse settings, as well as how they differ from what has been observed in these samples, is beneficial. Together, the meta-data analysis aided in the efficient

handling of large samples and the statistical study of variations in bacterial community composition due to environmental variables.

1.6 STUDY OBJECTIVES:

My main research goal is to use meta-data analysis to explore an environmental contamination phenomenon known as the "plastisphere." The bioinformatic analysis of whole meta-analysis pipeline revealed abundant amount of gene encodings of bacterial communities that colonised on plastic and organic particle by analysing the meta-database from existing research, amplicon sequencing of samples was performed by utilising the 16S rRNA gene sequence. There are three goals in investigating the environmental factors that influence bacterial assemblages: a) Examine the impact of geographic location on the microbial mix associated with diverse types of microplastics (marine plastic in seawater, sediments, and microplastic in freshwater). To precisely investigate spatial impacts, the samples from the experiment were compared to samples taken directly from the environment. b) Examine the microbial communities' characteristics and composition structure on various substrate types. The variety in bacterial community composition in response to diverse plastic types might reveal the colonisation factors. The biotic control samples (biofilm, wood, freshwater, seawater) were also expected to be examined, as was the comparison between the different bacterial populations on microplastic and other environmental samples. c) Determine the major variables influencing the microbiological composition of microplastics in marine environment based on various environmental circumstances (pH, salinity, plastic type etc.).

2. METHODOLOGY:

On the Orion cluster, the SRA toolkit and NCBI's e-utils were active in order to download sample sequences. To obtain related FASTQ data, the SRR numbers linked with Bioproject PRJNA for Seven investigations were sourced. The project folder was formed and then placed into a specific folder, after which a folder named Sequence folder was created and transferred into the folder, and the folder names were extracted from the paired-end files. The raw sequences were dumped into a "Raw" folder within each of these directories (as seen in step1 in appendix section).

2.1 Data Assemblage:

Because the project's major study focus is on processing and analysing the metadata of microorganisms adsorbed on microplastics, all of the projects data will come from publically available data on the NCBI site.

To bring the project data closer to the research theme, the data was primarily collected from academic articles which were focusing mainly on V4 region, microplastics, and microbes, and it was ultimately decided to use the data from the following Seven academic articles as the data basis for the project research analysis. These are the following project numbers used in this project:

- Bio Project PRJNA506548: Kesy et al.,2019 Spatial Environmental Heterogeneity Determines Young Biofilm Assemblages on Microplastics in Baltic Sea Mesocosms." Frontiers in Microbiology.
- Bio Project PRJNA242639: McCormick et al.,2016 "Microplastic in surface waters of urban rivers: concentration, sources, and associated bacterial assemblages". Ecosphere.

- Bio Project PRJNA338729: Oberbeckmann et al.,2018 “Environmental factors support the formation of specific bacterial assemblages on microplastics”. Frontiers in Microbiology.
- Bio Project PRJNA272679: De Tender et al.,2015 “Bacterial community profiling of plastic litter in the Belgian part of the North Sea”. Environmental science & technology.
- Bio Project PRJNA378706: Syranidou et al.,2019 “Biodegradation of mixture of plastic films by tailored marine consortia”. Journal of Hazardous Materials.
- Bio Project PRJNA283545: Oberbeckman et al.,2016 “Microbes on a Bottle: Substrate, Season and Geography Influence Community Composition of Microbes Colonizing Marine Plastic Debris”. Plos One.
- Bio Project PRJNA612500: Li et al.,2020 Impacts of microplastics exposure on mussel (*Mytilus edulis*) gut microbiota. Science of the Total Environment.

Study	Accession No	Number of samples	Sample Location	pH	Salinity	Region Targeted	Plastic Material	Platform
Luen Li_2020	PRJNA612500	144	Wimereux, France	7.6	32.2	V4	HDPE	Illumina MiSeq
McCormick_2014	PRJNA242639	16	North Shore Channel, Chicago, USA	7.2	.315	V4	Freshwater Biofilm	Illumina MiSeq
Tender_2015	PRJNA272679	109	North Sea, Belgium	8	33.5	V4	Sediment	Illumina MiSeq
Osborn_2016	PRJNA283545	93	North Sea, UK	7.4	32.4	V4	Polyethylene Terephthalate	Illumina MiSeq
Oberbeckmann_2018	PRJNA338729	149	Rostock, Germany	7.5-8.1	14.1	V4	Seawater Freshwater Wood PE PS Treated Wastewater	Illumina MiSeq
Schoch_2017	PRJNA378706	29	Greece	7.2	Unknown	V4	Polyethylene	Illumina MiSeq
Kesy_2019	PRJNA506548	300	Baltic Sea	8.1	4.4-9.0	V4	Seawater Wood PE PS PP	Illumina MiSeq

- ***Table1: Simpler version of the Meta-data table illustrating seven studies with 840 samples. Details like Project number, pH, Salinity, Region targeted, Plastic Material and Platform have been shown.***

2.2 FASTQ Collection:

To build a meta-database, seven studies were gathered from published research containing the search terms “Plastics”, and “Illumina Miseq”. The results were based on high-throughput sequencing of the 16S rRNA gene V4 region (paired end reads on Illumina MiSeq). Each of the research given looked at the development of microbial and biofilm assemblages in aquatic environments. In this research, the findings of the FASTQ file collection from reputable earlier research helped to classify the samples into groups based on specified factors.: isolation source, sample type, geological location, pH, and salinity, on the NCBI website and in previous research, extensive information about samples and bio-project IDs were recorded.

2.3 QIIME2 ANALYSIS:

The amplicon analysis was performed using QIIME2, which is a more sophisticated version of QIIME. Qiime2 is a relatively recent amplicon analysis method that is quite useful. We can create OTUs (at a 97 percent or any other threshold) as well as ASVs (Amplicon Sequencing Variants using DADA2). The table of Amplicon Sequence Variants (ASVs) was created in the first phase of the analysis sequence with the goal of identifying each specific amplicon sequence variation. A fake barcode was created to import data in the Earth Microbiome Project (EMP) format. A unique barcode was created for each of the sample readings. Forward readings were compiled into a single forward.fastq file to

summarise main data from several investigations. In the same way, all of the reverse readings were combined into a single reverse.fastq file. The folder "emp-paired-end-sequences" included all the FASTQ files created throughout the previous phases of data processing. The next step in the study was to import the "emp-paired-end- sequences" folder into the QIME2 platform. The sequences were demultiplexed using the DADA2 analysis techniques.

2.4 DADA2 ANALYSIS:

Raw amplicon sequencing data is converted into a database of precise amplicon sequence variations (ASVs) present in each sample in the DADA2. This technique helps us to focus on the thorough examination of microbial populations. DADA2 uses amplicon data to estimate precise amplicon sequence variations (ASVs), resolving biological differences as small as 1 or 2 nucleotides. After filtering, the DADA2 error model contains quality information that is disregarded by all other approaches. Most other approaches utilise abundance ranks if they use abundance at all, but the DADA2 error model uses quantitative abundances. Other techniques just count the mismatches, but the DADA2 error model highlights the variations between sequences, such as A->C. Instead of depending on prior datasets that may or may not represent the PCR and sequencing procedures used in your study, DADA2 may parameterize its error model from the data itself.

The DADA2 process begins with a collection of demultiplexed fastq files matching to amplicons-sequenced samples. In other words, DADA2 anticipates two distinct fastq files for each sample, one forward and one backward. The ASV table, like the DADA2 features previously discussed, records each sequence and its amount information to get microbiological categorization information with greater resolution than the OTU table produced by conventional clustering. All experiments were combined using QIIME2's merge-seqs commands, and then taxonomically categorised using a classifier trained on the SILVA database's full-

length 16S rRNA gene. Filters were used to eliminate mitochondria, chloroplasts, unidentified sequences at the kingdom level, and sequences with a cumulative abundance of ten or less.

Study_Sample	Number Reads Before QC	Reads After Dada2	Reads Resolved to Species	Reads Lost During 'Collation
Study 1	102754	37965	36998	967
Study 2	186637	141837	137703	4134
Study 3	176400	40661	40289	373
Study 4	101094	4770	4713	57
Study 5	86384	5369	5345	24
Study 6	19222	11203	2371	8832
Study 7	41176	26223	25845	378

Table 2: This table illustrates the total no. of reads before FastQC and the reads filtered after running DADA2. We ca also see the Reads Lost after the collation step was performed.

2.5 R Analysis:

The statistical analytic methods of alpha diversity, core-microbiome, Taxonomic-Bars and Heat trees will be utilised in this research to examine the microbiological diversity of microplastics and plastisphere. **Alpha diversity** is a comprehensive measure that reflects abundance and uniformity of microbial variety in each region or environment. Alpha diversity is primarily determined by two factors: the number of species (richness) and diversity (uniformity of individual distribution in the group).

Simpson Index: It's a type of index that's widely utilised in ecology. It represents the dominant species' position and 18 function in the community. When there are more dominant species in a community, the fraction of non-dominant species decreases. The Simpson index value is higher, indicating a lack of diversity in the community, and this index is adversely associated with other diversity indices.

Shannon Index: One of the indices used to calculate the microbial diversity in the sample. Shannon and the Simpson diversity index are two widely used alpha

diversity indices. The greater the Shannon value, the more diverse the community.

Core-Microbiome Heatmap: As a grid of coloured squares, a heatmap shows values for a primary variable of interest over two axis variables. Like a bar chart or histogram, the variables on the x,y axis are separated into ranges, and the colour of each cell reflects the value of the main core variable in the compatible cell range.

Taxa-Bar Plots: Summaries of taxonomic classifications. Stacked bar graphs depicting each taxon's average relative abundance at various taxonomic levels. Distinct coloured bars represent varied phyla (as stated by the key), and different tones reflect different orders and genera within the specified phylum in the Order and Genus level plots. Only phyla having a greater than 5% abundance in at least one sample of the exposure are displayed. "Other Bacteria" encompasses all other phyla.

Heat-Trees: Traditional taxonomical graphs are commonly used to visualise the massive data sets generated by new sequencing technologies of varied ecological systems, but the following techniques ignore the hierarchical nature of taxonomic classifications, and their dependence on colours for categories limits the number of taxa that can be distinguished. Heat trees can express a taxonomic hierarchy, but it's difficult to see how statistics are spread across the tree, particularly inside within taxa.

R studio will be used to evaluate the pre-processing dataset, and the findings in Alpha diversity, Core-microbiome heatmap, Heat trees and Taxonomic bars will be shown using a box plot, two-dimensional distribution chart, and histogram. The impacts of various circumstances (such as salinity, pH, and geographical location) on the microbial community composition were investigated using the standard of plastic type. The parameters of the samples were organised into a single metadata table for systematic review, and the table was then utilised in R statistical analysis in CSV format.

3. RESULTS:

3.1 ALPHA DIVERSITY:

This analysis will show plots comparing pH, Salinity, location with Plastic Types found in all different seven studies to analyse the microbial communities. Each graph will break down each index to show the influence of each variable on the microbial community more clearly. Because the diversity of a single sample cannot be quantified, it is only fair to compare the alpha diversity of the sample groups. The more species that contribute, the larger the alpha diversity depicted in the boxplot will be.

(a) As shown in figure 1. (a), this plot depicts the accumulated plastic types from seven different studies which is seen to be compared with diversity differences. Different collecting locations recognised the shapes, and the mean value was used in the quartile boxplot. Based on Pielou's evenness, Richness, Shannon, and Simpson alpha, PE and PS samples had somewhat greater richness and total diversity. The Simpson index for each set of samples was considerably high (almost 1.0), indicating that each group of samples had a lot of diversity.

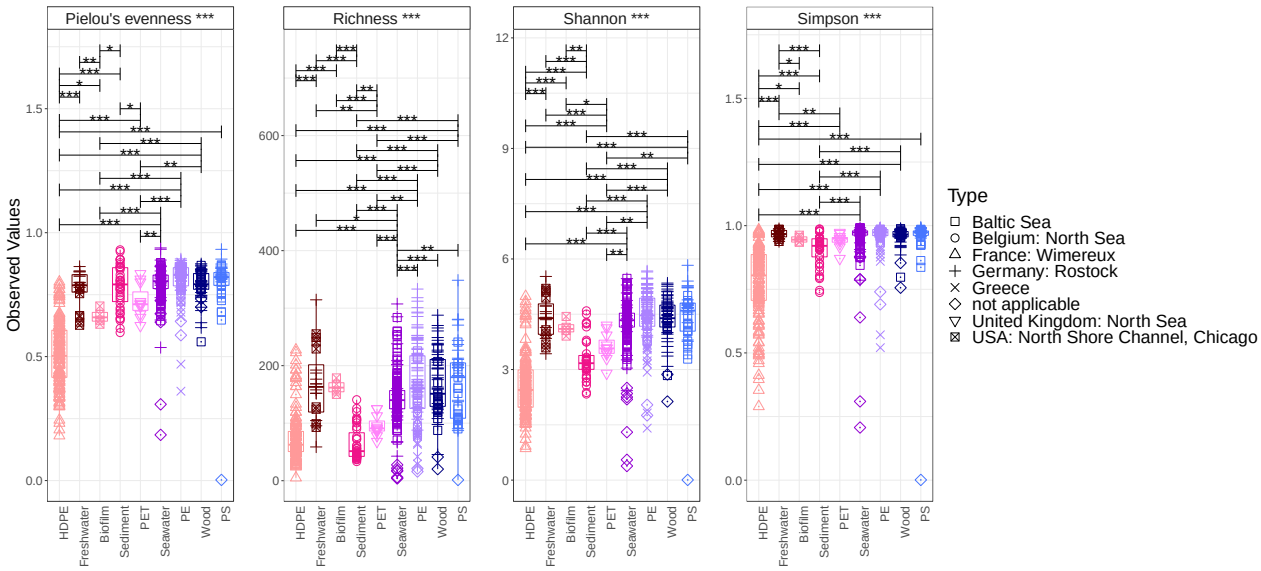


Figure 1: (a) Alpha Diversity plot depicts the Plastic Type v/s Location. It Species Richness with the corresponding colour legend for the treatment groups. Pair-wise ANOVA P-values are displayed P < 0.05*, P < 0.01**, P < 0.001*.

B) As shown in Figure1.(b) this plot depicts effect of pH on the microbial community diversity on various types of plastic materials. To demonstrate the distribution of microbial variety on different types of plastics, the pH value would be split into six factors. Looking at Simpson Index we can surely say that according to the P value (P < 0.05*, P < 0.01**, P < 0.001***) the microbial diversity in sample groups of various plastic materials is only marginally significant. The sample distribution is very consistent within a single sample group, therefore the variation between individual samples is minimal. The limited microbial diversity is reflected in the Simpson index value of almost one.

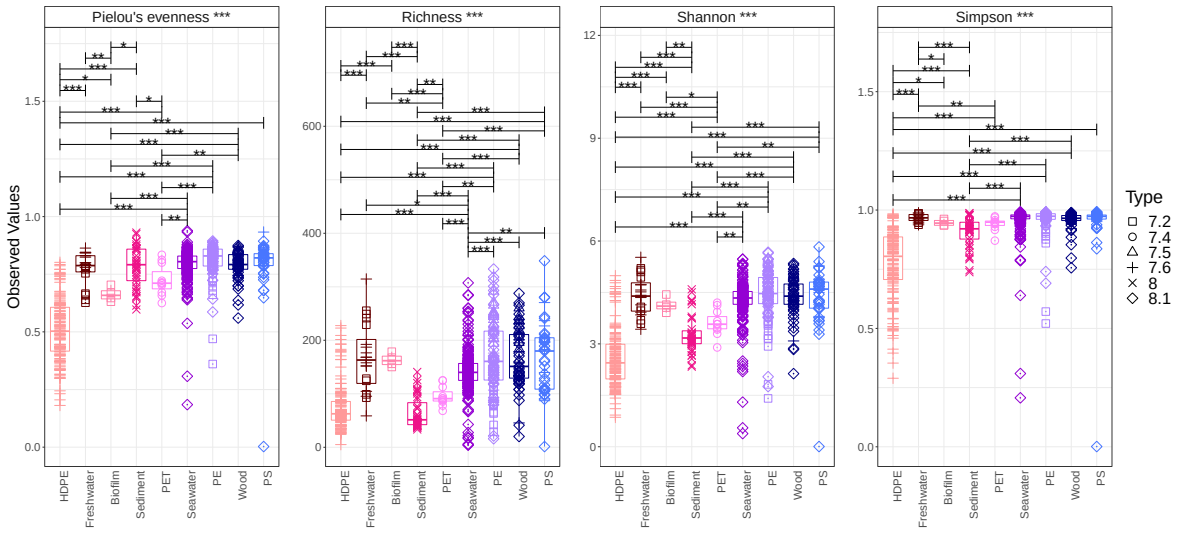


Figure 1: (b) Alpha Diversity plot depicts the Plastic Type v/s pH.

c) As shown in figure 1. (c), this plot depicts the effects of salinity which was divided into six variables to compare the diversity of microbial communities on different plastic types. P value ($P < 0.05$*, $P < 0.01$**, $P < 0.001$***) indicates that the difference in microbial diversity among different materials of plastic samples is substantial, according to the richness index. Because the sample distribution is not intensive, the variation between individual samples is significant. Because of the abundance of microbial species, the samples have a wide range of variety.

The variations in the diversity of the microbial communities indicated in the project data are considerable, according to the study of the above three graphs, and the uniformity and richness are also pretty big. Because the information provided by each image is similar. As a result, external environmental variables have little impact on the microbial population on various plastic materials.

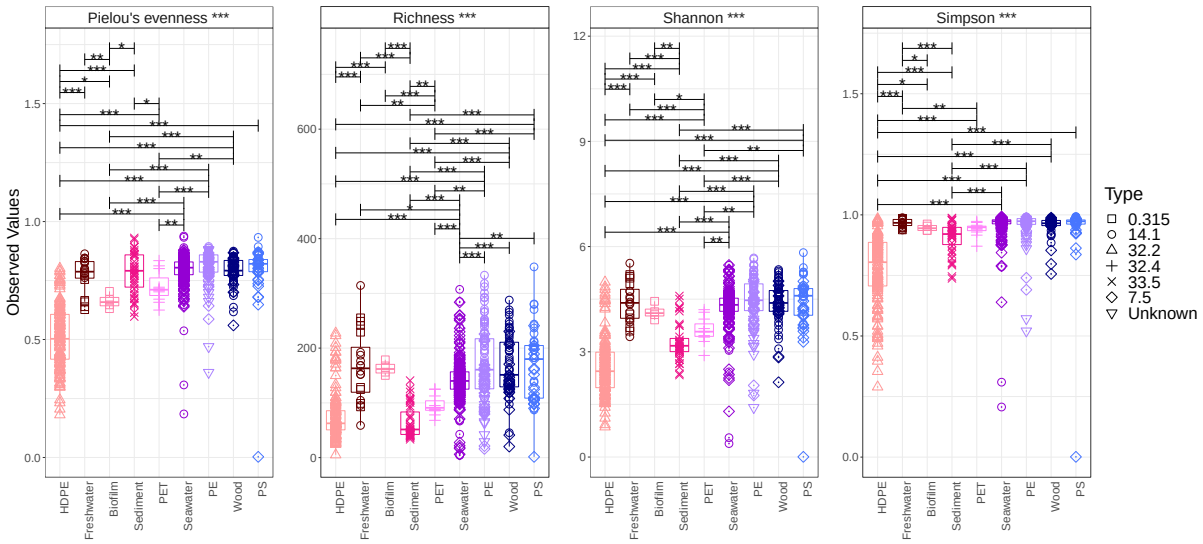


Figure1: (c) Alpha Diversity plot depicts the Plastic Type v/s Salinity.

3.2 Taxa-Bars:

Each population's top-25 most prevalent genera are displayed. Taxonomic classification at a finer level (the ASV level). The initial insight into the similarities and differences of microbial diversity at the family and phylum level came from observational analyses of the top-25 most prevalent genera between the plastic types.

A) As shown in figure 2 (a) the plastic type of phylum taxa-bars are being looked at. In Phylum level, Proteobacteria (Approx 50%) is found in abundance in all the plastic samples. Bacteroidota (Approx 20%) can also be seen in abundance after Proteobacteria, and can be said as the second highest phylum type in this observation. Actinobacteriota can be considered as the third highest with approximately 10% of the total microbial diversity. These three bacteria accounts for the majority of microorganisms in all plastic materials, with the distribution of other bacteria being fairly even. All plastics produced from seawater include the most varieties of microorganisms, indicating that seawater is an ideal habitat for the development of microbial communities.

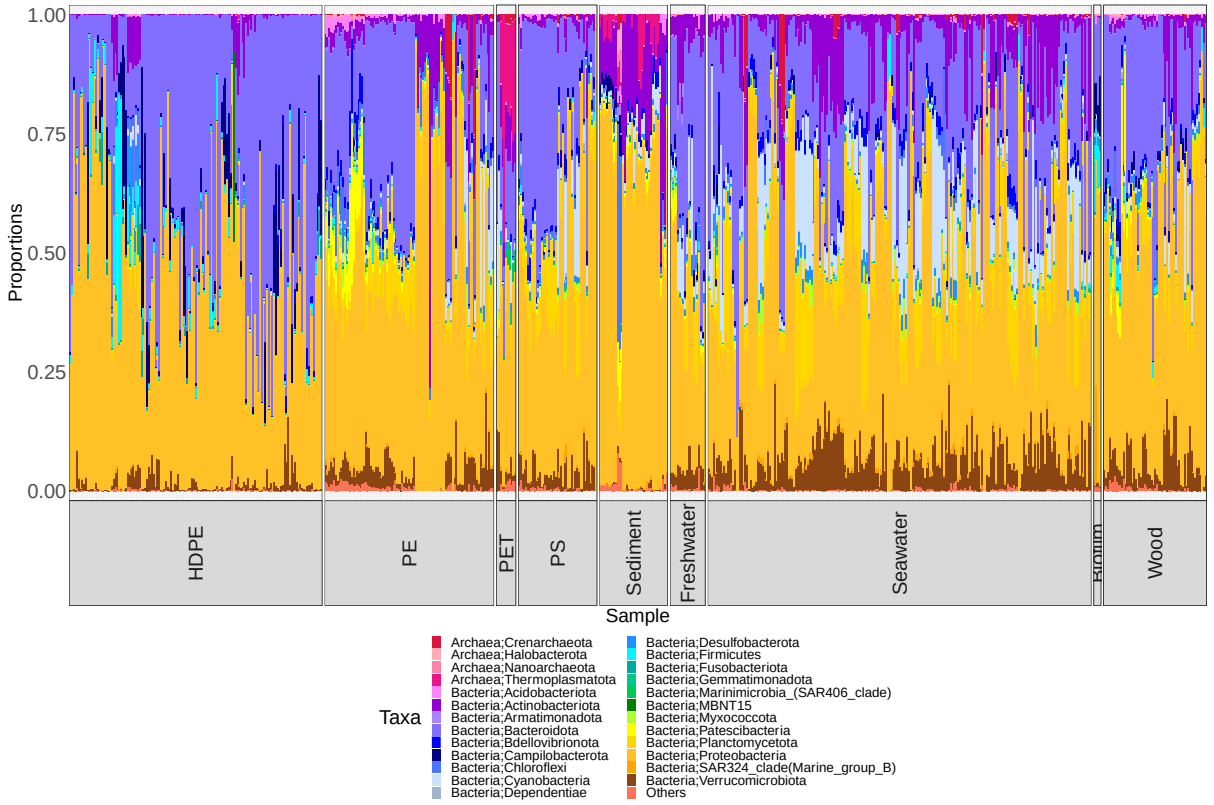


Figure 2 (a) Plastic types of Phylum Taxa Bars.

b) As shown in figure 2 (b) the plastic type of family taxa-bars are being looked at. *Flavobacteriaceae* family can be seen in abundance consisting of 20% of the whole diversity. Rhodobacteraceae can be seen as the second most seen diverse family. Nitrincolaceae is third most diverse family in the whole graph. The distribution of other bacteria is very consistent, and these three species of bacteria account for the biggest share of all plastic types. All plastics produced from seawater include the most varieties of microorganisms, indicating that seawater is an ideal habitat for the development of microbial communities.

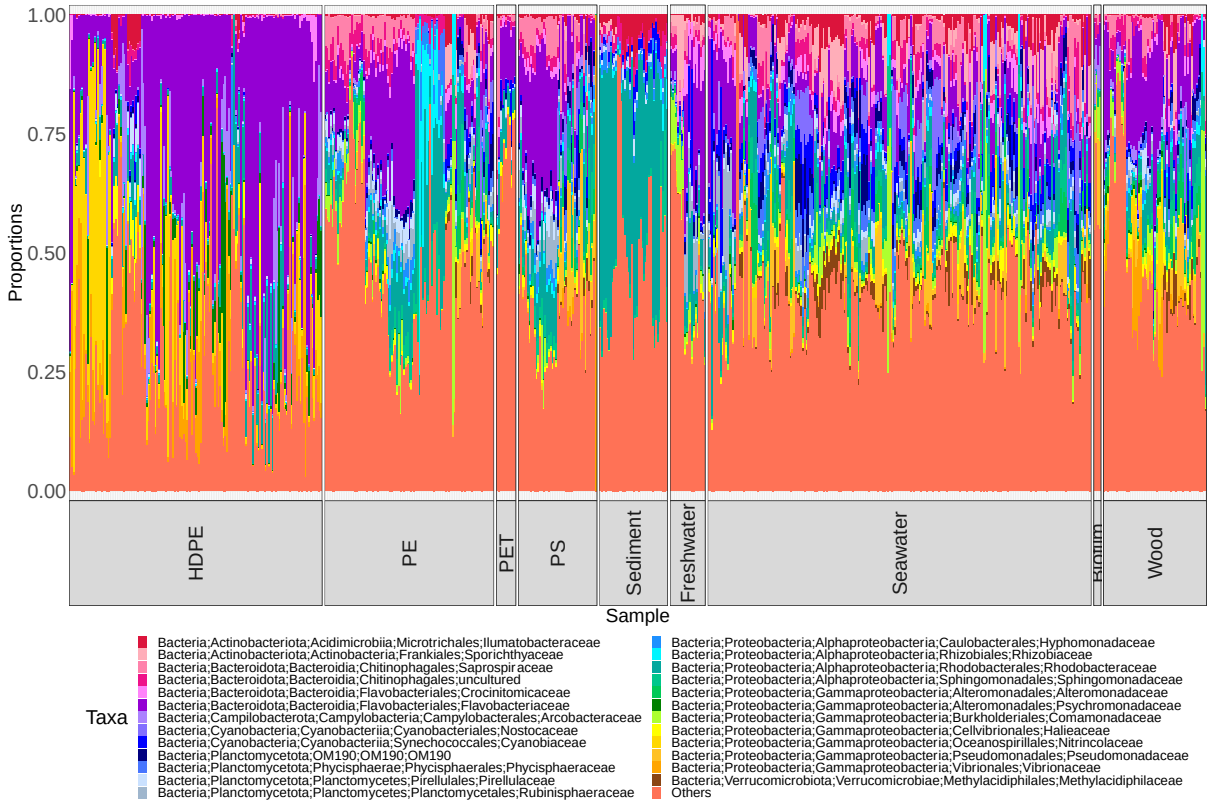


Figure 2 (b) Plastic types of Family Taxa Bars.

3.3 Heat Trees:

The taxonomic map is a legend that shows the distinguishing taxa's identities. The differential heat trees are informed by this mythology. The microbiota changes from different plastic materials are shown in differential heat trees. The branches where they are upregulated are coloured in accordance with the tree's corresponding classifications. The legend for the nodes is as follows: Right side 'Number of ASVs,' where the diameter reflects the number of ASVs corresponding to the nodes on the heat trees, left side 'Log2 ratio' of the median proportions of the taxa corresponding to the colour scale, right side 'Number of

ASVs,' where the diameter reflects the number of ASVs corresponding to the nodes.

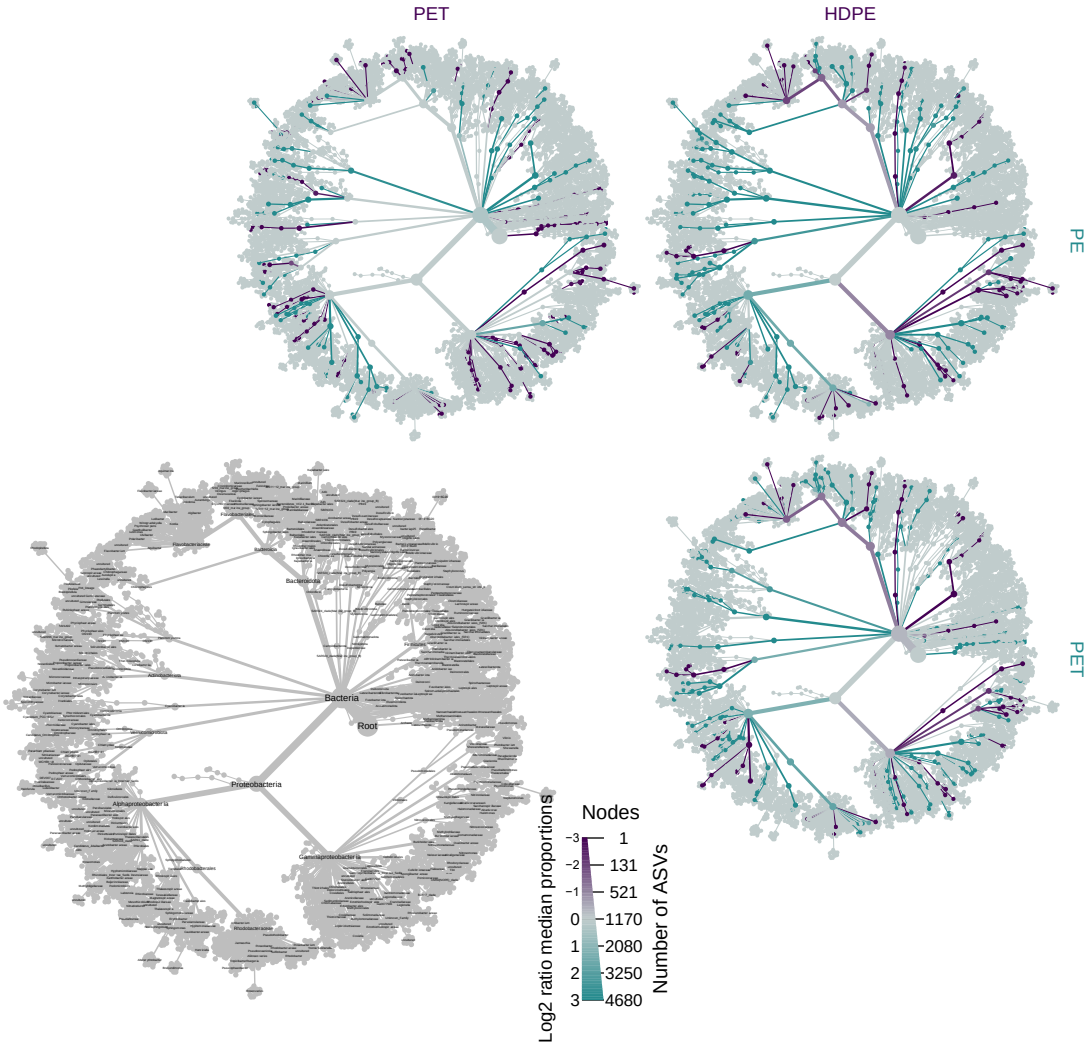


Figure 3 (a) Depicts 3 categories of plastics i.e., PE, PET and HDPE.

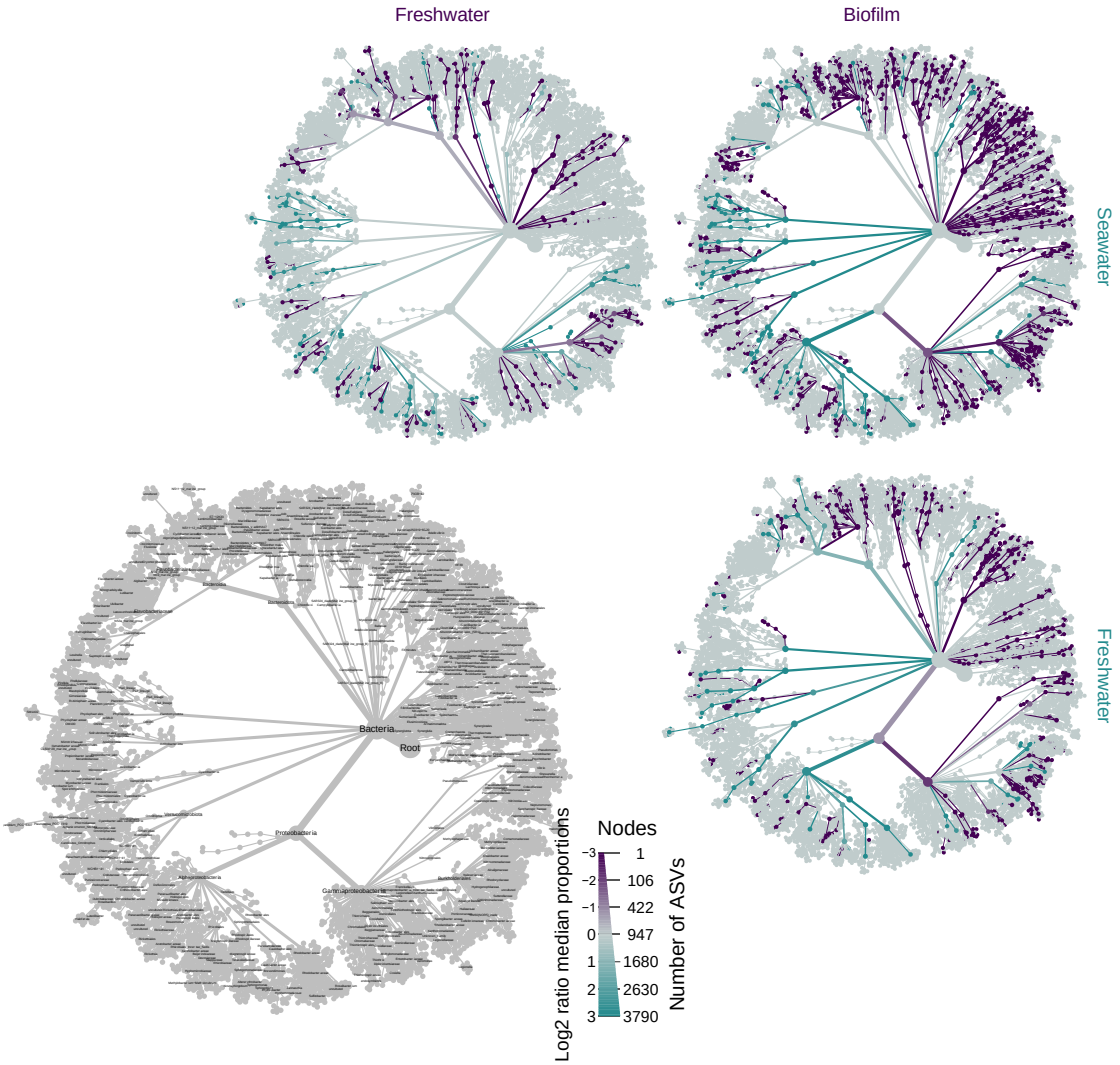


Figure 3 (b) Depicts the next three categories of plastics i.e., Freshwater, Seawater and Biofilm.

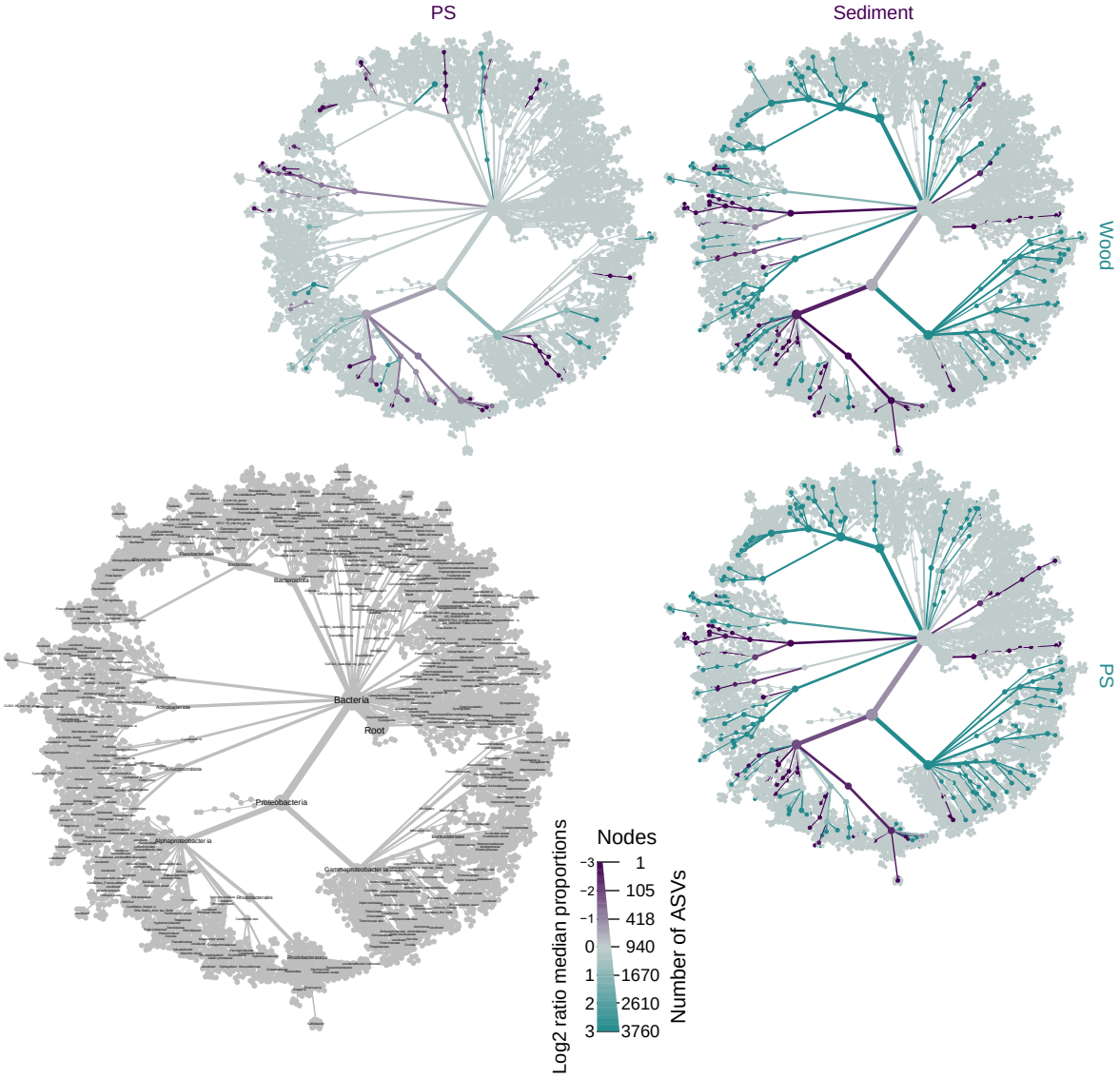


Figure 3 (c) Depicts the next three categories of plastics i.e., Wood, Sediment and PS.

The above figures 3 (a), (b) and (c) depicts that Heat trees may plot several trees per graph and can present up to three metrics in a taxonomic context. The size and colour of text, nodes, and edges may all be automatically translated to arbitrary numbers, allowing for the quantitative display of numerous statistics at the same time.

Observations of the heat tree show (Figure 3b) that there are most number of branches that are highlighted as comparison between Biofilm and Seawater outside of those selected by the subset analysis. The other two comparisons between HDPE, PE and Sediment, wood is not as upregulated as Biofilm and seawater obtained plastic types. The log-2 ratio of median proportions of readings recorded at each body location is shown by the colour of each taxon. Only significant changes are shown in colour. Each node represents an ASV taxon, and the edges indicate where it belongs in the larger taxonomic hierarchy.

3.4 Core-Microbiome Heatmap:

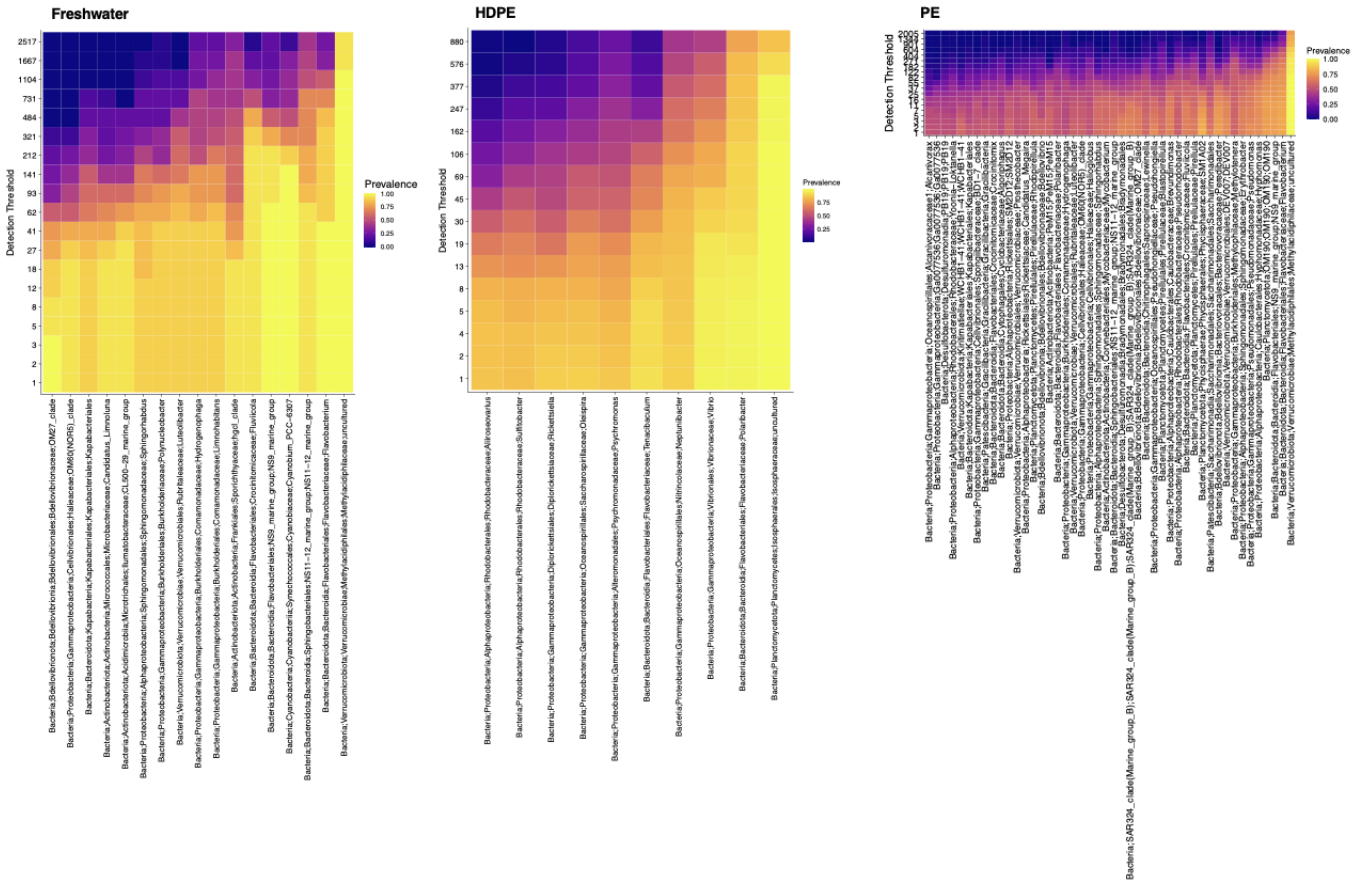


Figure 4 (a) Depicts the core-heatmap analysis of three plastic types i.e., freshwater, HDPE and PE.

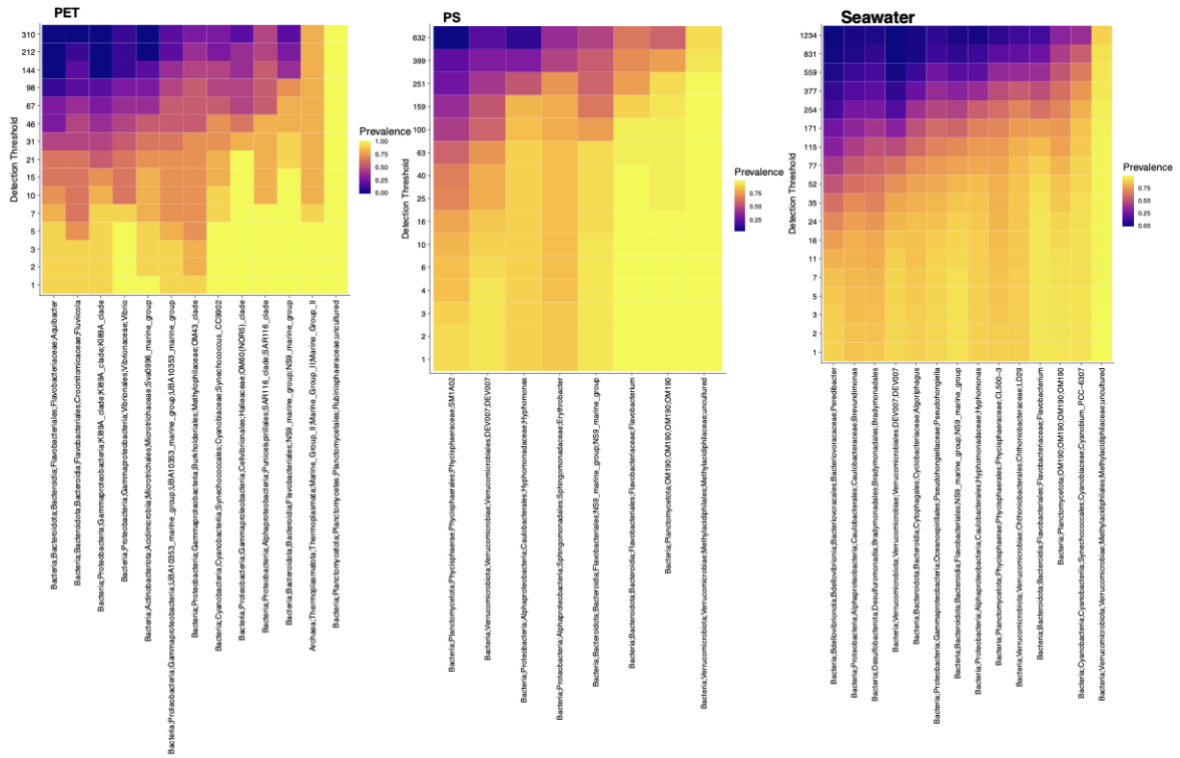


Figure 4 (b) Depicts the core-heatmap analysis of plastic types i.e., PET, PS and Seawater.

Heatmaps are ideal for visualising microarray data or data from high-throughput sequencing investigations like microbiome research. They're basically false colour pictures in which cells in a matrix with high relative values are coloured differently than cells with low relative values. Heatmaps can be as basic as two-sided colour blocks with lists, or they might incorporate information about hierarchical clustering and/or values of other variables of interest. R has several tools for creating and annotating heatmaps, which is great.

When binned at the Genus level, a heatmap of OTU abundances from each sample is created. The core microbiome revealed the microbial diversity common to all combinations of plastic types, where genera had at least 85% frequency. In the heat map figures, genera are classified by their abundance, with the low abundant prevalent genera at the top and most abundant prevalent genera at the bottom. The highest amount of diversity can be seen in Polyethylene with

Methylacidiphilobacterium being in abundance and flavobacterium being the second most abundant.

4. DISCUSSION:

I wanted to look at the environmental effect of microplastics as well as the influence of microbial communities on microplastics in this study. According to the results of the analysis, changing location enhances the environmental effect on the microbial community, but environmental variables (pH, salinity) have no substantial impact on the microbial community's composition and impact. Under extreme environmental conditions (Antarctic, Arctic, and submarine volcanoes), it can have a significant impact on the formation and composition of microbial communities (Li et al., 2014).

The data analysis in the findings section, however, reveals that the samples used were from typical marine habitats, such as sediments and microplastics floating in shallow waters, which are not harsh ecosystems under such conditions.

4.1 Plastisphere Diversification of PE and Biofilm:

The possibly dangerous Proteobacterium (Pseudomonadales and Alteromonadales) and Flavobacterium were discovered in the PS, PE, and biofilm plastic types, according to the analysis done in core-microbiome heatmap. In the saltwater, Pseudomonadales were found to be more concentrated in Biofilms and PE, whereas Flavobacterium was found to be abundant in most of the plastic types used in this research. It demonstrates that the presence of these microbes may play a role in populating the microplastic via selection preference. The PE and Biofilms had the highest levels of pathogenic colonisation. But, keeping in mind that these Proteobacteriums can be highly dangerous for human health’s as

Pseudomonadales can cause disease like pneumonia, chronic opportunistic infections.

4.2 Bacteria Community's Complex Carbon Degradation of Microplastics:

Microplastic biodegradation is a process that converts organic carbon into biomass and biogas, either partially or completely. As a result, microplastic breakdown is linked to the activity of the bacteria population that consumes microplastic as a carbon source. Various circumstances in freshwater and seawater, allow for different types of deterioration. Microplastics can be used as a single carbon source by Pseudomonadales and Alteromonadales found in PE and Biofilm. This degradation occurs alongside the chemical degradation through the weakening of the polymers, evidenced by the molecular changes and roughness of the microplastics (Jacquin et al., 2019).

5. CONCLUSION:

In this study, I re-analysed and combined 7 amplicon sequencing datasets from various Plastisphere studies. The study shows that intrinsic factors have a major role in determining the bacteria community structure in freshwater and seawater, based on diversity analysis and taxonomic annotation. The major influence was sample collecting site, while the substance of microplastics was found as the secondary component. When compared to organic matter, Pseudomonadales and Altermonadales with a greater prevalence of plastic material and a high proportion of Flavobacterium in polyethylene explain the techniques of selection for bacteria community formation under varied environmental stresses. The alpha diversity, on the other hand, denotes the relative correlation of community structure in marine life, and the link across ecosystems causes microplastic to be transported with potentially dangerous bacteria. As a result, researchers intend to look into the relationship between pathogenic colonisation and carbon breakdown on microplastics. Weakening of polymers in the aquatic environment generates roughness, and molecular alterations on the surface might inspire plastisphere sculpting.

References:

Thiel, M., Luna-Jorquera, G., álvarez-Varas, R., Gallardo, C., Hinojosa, I. A., Luna, N., et al. (2018). Impacts of marine plastic pollution from continental coasts to subtropical gyres-fish, seabirds, and other vertebrates in the SE Pacific. *Front. Mar. Sci.* 5:238. doi: 10.3389/fmars.2018.00238

Smith, M., Love, D. C., Rochman, C. M., and Neff, R. A. (2018). Microplastics in seafood and the implications for human health. *Curr. Environ. Health Rep.* 5, 375–386. doi: 10.1007/s40572-018-0206-z

Shen, M., Zhang, Y., Zhu, Y., Song, B., Zeng, G., Hu, D., et al. (2019). Recent advances in toxicological research of nanoplastics in the environment: a review. *Environ. Pollut.* 252, 511–521. doi: 10.1016/j.envpol.2019.05.102

Keswani, A., Oliver, D. M., Gutierrez, T., and Quilliam, R. S. (2016). Microbial hitchhikers on marine plastic debris: human exposure risks at bathing waters and beach environments. *Mar. Environ. Res.* 118, 10–19. doi: 10.1016/j.marenvres.2016.04.006

Horton, A. A., Walton, A., Spurgeon, D. J., Lahive, E., and Svendsen, C. (2017). Microplastics in freshwater and terrestrial environments: evaluating the current understanding to identify the knowledge gaps and future research priorities. *Sci. Total Environ.* 586, 127–141. doi: 10.1016/j.scitotenv.2017.01.190

Allen, S., Allen, D., Phoenix, V. R., Le Roux, G., Durántez Jiménez, P., Simonneau, A., et al. (2019). Atmospheric transport and deposition of

microplastics in a remote mountain catchment. *Nat. Geosci*. 12, 339–344. doi: 10.1038/s41561-019-0335-5

Ambrosini, R., Azzoni, R. S., Pittino, F., Diolaiuti, G., Franzetti, A., and Parolini, M. (2019). First evidence of microplastic contamination in the supraglacial debris of an alpine glacier. *Environ. Pollut.* 253, 297–301. doi: 10.1016/j.envpol.2019.07.005

Nauendorf, A., Krause, S., Bigalke, N. K., Gorb, E. V., Gorb, S. N., Haeckel, M., et al. (2016). Microbial colonization and degradation of polyethylene and biodegradable plastic bags in temperate fine-grained organic-rich marine sediments. *Mar. Pollut. Bull.* 103, 168–178. doi: 10.1016/j.marpolbul.2015.12.024

M.R. Gregory, Plastic 'scrubbers' in hand cleansers: a further (and minor) source for marine pollution identified
Mar. Pollut. Bull., 32 (1996), pp. 867-871

K. Duis, A. Coors
Microplastics in the aquatic and terrestrial environment: sources (with a specific focus on personal care products), fate and effects Environ. Sci. Eur., 28 (2016), p. 2

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME journal, 11(12), 2639-2643.

Zettler ER, Mincer TJ, Amaral-Zettler LA. Life in the "Plasti- sphere": Microbial communities on plastic marine debris. Environ Sci Technol. 2013;47:7137

Paço, Ana; Jacinto, Jéssica; Costa, João Pinto da; Santos, Patrícia S. M.; Vitorino, Rui; Duarte, Armando C.; Rocha-Santos, Teresa (2019-03-04).

Thomas, Russell (14 June 2021). "Plastic rafting: the invasive species hitching a ride on ocean litter". *The Guardian*.

Quero, G. M., & Luna, G. M. (2017). Surfing and dining on the “plastisphere”: Microbial life on plastic marine debris. *Advances in Oceanography & Limnology*, *8*(2), 199–207.

Li,S.J.,Hua,Z.S.,Huang,L.N.,Li,J.,Shi,S.H.,Chen,L.X.,...&Shu, W. S. (2014). Microbial communities evolve faster in extreme environments. Scientific reports, 4, 6205.

Jacquin, J., Cheng, J., Odobel, C., Pandin, C., Conan, P., Pujo-Pay, M., ... & Ghiglione, J. F. (2019). Microbial ecotoxicology of marine plastic debris: a review on colonization and biodegradation by the "plastisphere *Frontiers in microbiology*, *10*, 865.

APPENDIX:

The linux commands were provided by my supervisor Dr. Umer Ijaz.

How to download data using SRI Toolkit

Imagine, I go to BioProject website and I search for a particular keyword:

https://www.ncbi.nlm.nih.gov/bioproject/?term=16S+rRNA+water+microbiome

SRA Handbook:
https://www.ncbi.nlm.nih.gov/books/NBK242621/

Inorder to download sequences, we need to worry about two things:

a) Bioproject ID
b) SRA IDs (is for samples)

Secondly, we need to have two software suite installed on the cluster/local computer (MacBook)

NCBI's e-utils (edirect)
https://www.ncbi.nlm.nih.gov/books/NBK25500/

SRA toolkit
https://www.ncbi.nlm.nih.gov/sra/docs/sradownload/

On Orion cluster, if you want to enable both software on your account, you need to type these lines (to make them available in your path).

```
export PATH=/home/opt/sratoolkit.2.9.0-centos_linux64/bin:$PATH
export PATH=/home/opt/edirect:$PATH
```

SRA location:
https://www.ncbi.nlm.nih.gov/sra

Step 1: Get all the SRR numbers associated with a bioproject PRJNA*

```
esearch -db sra -query PRJNA302804 | efetch --format runinfo |cut -d "," -f 1 > SRR.numbers
```

Step 2: Retain only the SRR numbers in the files

```
awk '/SRR/' SRR.numbers > SRR.numbers.filtered
```

QIIME 2 WORKFLOW:

First thing to do: We have all the sequences available in a given folder

```
mkdir shubhra
cd shubhra
mkdir sequences
cd sequences
cp /home/opt/tutorials/Raw/*.fastq .
ls
```

Step 1: We are going to organize our data in such a manner that for every sample we have the folder name extracted from the paired-end files, and we are going to dump the raw sequences in a "Raw" folder:

```
[MScBioinf@becker ~/shubhra/sequences]$ for i in $(awk -F"_" '{print $1}' <(ls *.fastq) | sort | uniq); do mkdir $i; mkdir $i/Raw; mv $i*.fastq $i/Raw/.; done

[MScBioinf@becker ~/shubhra/sequences]$ cd ..
[MScBioinf@becker ~/ Step 1: We have all the sequences available in a given folder

mkdir shubhra
cd shubhra
mkdir sequences
cd sequences
cp /home/opt/tutorials/Raw/*.fastq .
ls
```

Step 1: We are going to organize our data in such a manner that for every sample we have the folder name extracted from the paired-end files, and we are going to dump the raw sequences in a "Raw" folder:

```
[MScBioinf@becker ~/shubhra/sequences]$ for i in $(awk -F"_" '{print $1}' <(ls *.fastq) | sort | uniq); do mkdir $i; mkdir $i/Raw; mv $i*.fastq $i/Raw/.; done
```

```
[MScBioinf@becker ~/shubhra/sequences]$ cd ..
[MScBioinf@becker ~/shubhra]$ mkdir qiime2_tutorial
[MScBioinf@becker ~/shubhra]$ cd qiime2_tutorial
[MScBioinf@becker ~/shubhra /qiime2_tutorial]$
```

There are different ways in which we can import data to qiime2: https://docs.qiime2.org/2020.2/tutorials/importing/

```
[MScBioinf@becker ~/shubhra /sequences]$ d="/home/eng/MScBioinf/shubhra/sequences/";
[MScBioinf@becker ~/shubhra/sequences]$ cd ../qiime2_tutorial
```

Next step is to generate fictitious barcodes required to import data in Earth Microbiome Project (EMP) format (consult http://userweb.eng.gla.ac.uk/umer.ijaz/bioinformatics/oneliners.html#PERLOL on how I have written one-liner in perl to generate fictitious barcodes):

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ t=$(ls $d | wc -l);
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ paste <(ls $d) <(perl -le 'sub p{my $l=pop @_;unless(@_){return map [$_],@$l;}return map { my $ll=$_; map [@$ll,$_],@$l} p(@_);} @a=[A,C,G,T]; print join("", @$_) for p(@a,@a,@a,@a,@a,@a,@a,@a);' | awk -v k=$t 'NR<=k{print}') | awk 'BEGIN{print "sample-id\tbarcode-sequence\n#q2:types\tcategorical"}1' > sample_metadata.tsv
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ cat sample_metadata.tsv
sample-id	barcode-sequence
#q2:types	categorical
109-2	AAAAAAAA
1-1	AAAAAAAC
110-2	AAAAAAAG
113-2	AAAAAAAT
114-2	AAAAAACA
115-2	AAAAAACC
117-2	AAAAAACG
119-2	AAAAAACT
120-2	AAAAAAGA
126-2	AAAAAAGC
128-2	AAAAAAGG
130-2	AAAAAAGT
13-1	AAAAAATA
132-2	AAAAAATC
20-1	AAAAAATG
27-1	AAAAAATT
32-1	AAAAACAA
38-1	AAAAACAC
45-1	AAAAACAG
51-1	AAAAACAT
56-1	AAAAACCA
62-1	AAAAACCC
68-1	AAAAACCG
7-1	AAAAACCT
```

Step 2: Generate barcodes for each read using the file as above

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ (for i in $(ls $d); do bc=$(awk -v k=$i '$1==k{print $2}' sample_metadata.tsv); bioawk -cfastx -v k=$bc '{print "@"$1" "$4"\n"k"\n+";for(i=0;i<length(k);i++){printf "#"};printf "\n"}' $d/$i/Raw/*_R1_*.fastq ; done) > barcodes.fastq
```

Essentially, we are extracting the read headers from all the forward FASTQ files, and we assign the barcodes generated from sample_metadata.tsv file to those headers

```
[MScBioinf@becker        ~/shubhra/qiime2_tutorial]$        head
barcodes.fastq
@M01359:18:000000000-A5HVT:1:1101:15648:3435 1:N:0:109
AAAAAAAA
+
########
@M01359:18:000000000-A5HVT:1:1101:15642:3453 1:N:0:109
AAAAAAAA
+
########
@M01359:18:000000000-A5HVT:1:1101:22693:3963 1:N:0:109
AAAAAAAA
[MScBioinf@becker        ~/shubhra/qiime2_tutorial]$        tail
barcodes.fastq
+
########
@M01359:18:000000000-A5HVT:1:2114:15029:28668 1:N:0:68
AAAAACCG
+
########
@M01359:18:000000000-A5HVT:1:1106:13378:23600 1:N:0:7
AAAAACCT
+
########
```

Step 3: Collate all the forward reads from all the folders together in a single forward.fastq file

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ (for i in $(ls
$d); do cat $d/$i/Raw/*_R1_*.fastq ; done) > forward.fastq
```

Step 4: Assemble all the reverse reads from all the folders together in a single reverse.fastq file

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ (for i in $(ls
$d); do cat $d/$i/Raw/*_R2_*.fastq ; done) > reverse.fastq
```

See if the numbers match

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ ls
barcodes.fastq           forward.fastq           reverse.fastq
sample_metadata.tsv
[MScBioinf@becker  ~/shubhra/qiime2_tutorial]$  bioawk  -cfastx
'END{print NR}' forward.fastq
554815
[MScBioinf@becker  ~/shubhra/qiime2_tutorial]$  bioawk  -cfastx
'END{print NR}' reverse.fastq
```

```
554815
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ bioawk -cfastx 'END{print NR}' barcodes.fastq
554815
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$
```

Step 5: Zip all the FASTQ files and move them to emp-paired-end-sequences folder

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ gzip *.fastq
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ mkdir emp-paired-end-sequences; mv *.gz emp-paired-end-sequences/.
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ ls
emp-paired-end-sequences  sample_metadata.tsv
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$
```

Next, we enable Qiime2 on the Orion cluster

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ export PATH=/home/opt/miniconda2/bin:$PATH
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ source activate qiime2-2019.7
```

Step 6: Import the sequences to Qiime2

```
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime tools import --type EMPPairedEndSequences --input-path emp-paired-end-sequences --output-path emp-paired-end-sequences.qza
```

Step 7: Demultiplex the sequences in Qiime2

```
(qiime2-2019.7) [MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime demux emp-paired --p-no-golay-error-correction --i-seqs emp-paired-end-sequences.qza --m-barcodes-file sample_metadata.tsv --m-barcodes-column barcode-sequence --o-per-sample-sequences demux.qza --o-error-correction-details demux-details.qza
```

Step 8: Depends on the quality of your run, we want to fine tune Dada2 algorithm by specifying the thresholds

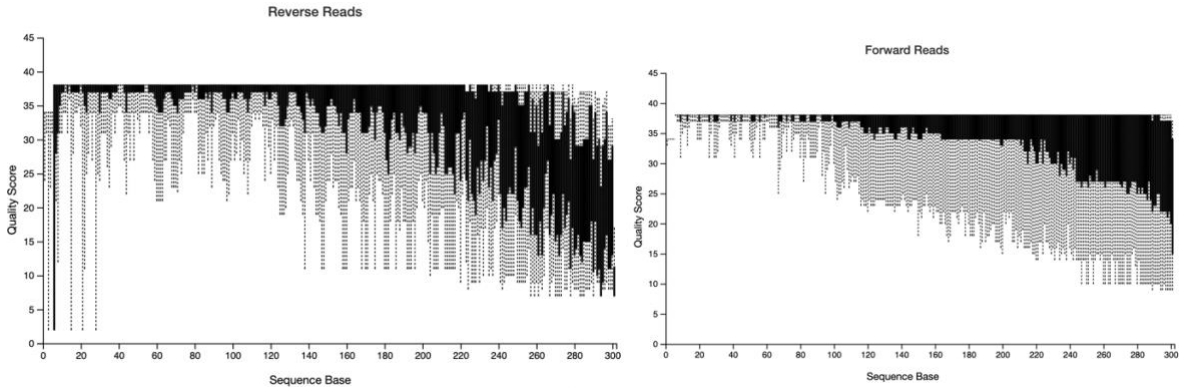
```
(qiime2-2019.7) [MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime demux summarize --i-data ./demux.qza --o-visualization ./demux.qzv

(qiime2-2019.7) [MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime tools export --input-path demux.qzv --output-path output
```

Next download the file to your local computer

scp MScBioinf@becker.eng.gla.ac.uk:~/shubhra/qiime2_tutorial/demux.qzv .

Next drag and drop the file on the Qiime2 viewer https://view.qiime2.org and manually figure out the thresholds, i.e., where the quality drops down significantly



Step 9: Run Dada2 algorithm

```
(qiime2-2019.7) [MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime dada2 denoise-paired --i-demultiplexed-seqs demux.qza --p-trim-left-f 0 --p-trim-left-r 0 --p-trunc-len-f 240 --p-trunc-len-r 200 --p-n-threads 0 --o-table table.qza --o-representative-sequences rep-seqs.qza --o-denoising-stats denoising-stats.qza –verbose
```

Step 9: Run Dada2 algorithm (record in your notes)

A)

```
(qiime2-2019.7)
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime dada2 denoise-paired --i-demultiplexed-seqs demux.qza --p-trim-left-f 0 --p-trim-left-r 0 --p-trunc-len-f 230 --p-trunc-len-r 240 --p-n-threads 0 --o-table table.qza --o-representative-sequences rep-seqs.qza --o-denoising-stats denoising-stats.qza --verbose
```

B)

```
qiime metadata tabulate --m-input-file denoising-stats.qza --o-visualization stats-dada2.qzv
```

C)

```
qiime feature-table summarize --i-table table.qza --o-visualization table.qzv
```

D)

qiime feature-table tabulate-seqs --i-data rep-seqs.qza --o-visualization rep-seqs.qzv

table.qza : Created the feature-table.biom file to represent ASVs table.

rep-seqs.qza: Created the dna-sequences.fasta file to represent ASV sequences.

Step 10: Export all the files that Qiime2 generated

```
(qiime2-2019.7)
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime tools
export --input-path table.qza --output-path output
Exported table.qza as BIOMV210DirFmt to directory output

(qiime2-2019.7)
[MScBioinf@becker ~/shubhra/qiime2_tutorial]$ qiime tools
export --input-path rep-seqs.qza --output-path output
Exported        rep-seqs.qza as DNASequencesDirectoryFormat to
directory output

MOVE INTO THE OUTPUT DIRECTORY

cd output

(qiime2-2019.7)
[MScBioinf@becker ~/shubhra/qiime2_tutorial/output]$ biom conv
ert -i feature-table.biom -o feature-table.tsv --to-tsv
```

Step 11: Generate taxonomy for these ASVs using Bayesian Lowest Common Ancestor Approach (BLCA) and greengenes database

Step 11A: Enable BLCA

Deactivate qiime if it activated

source deactivate

```
[MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$

source /home/opt/BLCA/enable_BLCA.sh
```

Step 11B: Run BLCA

Option 1 (SILVA138):

```
(python3) [MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$
```

```
2.blca_main.py                                    -i dna-sequences.fasta -q
/home/opt/qiime2_databases/silva-138-99-seqs/dna-sequences.fasta -r
/home/opt/qiime2_databases/silva-138-99-tax/taxonomy_BLCA.txt
```

Step 12: Enable Qiime2

```
(python3)
[MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$ source
deactivate
[MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$ export
PATH=/home/opt/miniconda2/bin:$PATH
[MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$ source
activate qiime2-2019.7
```

The next steps are to creat the tax biom file. You need to do this in the output ofolder of each project. Use one screen and just cycle through.

Step 13: Convert the BLCA format to regular qiime2 taxonomy format

```
(python3) [MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$

awk                              -F"\t"                              'BEGIN{print
"FeatureID\tTaxon\tConfidence"}{gsub("superkingdom:","D_0__",$2);gsub("
phylum:","D_1__",$2);gsub("class:","D_2__",$2);gsub("order:","D_3__",$2
);gsub("family:","D_4__",$2);gsub("genus:","D_5__",$2);gsub("species:",
"D_6__",$2);gsub(";$|\t$","",$2);gsub(";"," ;",$2);gsub(" ;[0-9]+(\.[0-
9]+)?      ;",";",$2);gsub("Unclassified","Unassigned\t100.0",$2);gsub("
+;","\t",$2);print                                                $1"\t"$2}' dna-
sequences.fasta.blca.out > taxonomy_silva_BLCA.tsv

If you haven't already change the biom file to tsv format
(qiime2-2019.7)
[MScBioinf@becker ~/SHUBHRA/qiime2_tutorial/output]$ biom conv
ert -i feature-table.biom -o feature-table.tsv --to-tsv


(qiime2-2019.7)
[MScBioinf@becker /shared5/AD_Fanatics/qiime2_V1_V2/output]$

biom add-metadata                  -i feature-table.tsv                  -
o feature_w_tax_silva_BLCA.biom                  --observation-metadata-
fp taxonomy_silva_BLCA.tsv                              --observation-
header FeatureID,taxonomy,Confidence    --sc-separated taxonomy --float-
fields Confidence
```

Step 14:

Using cyberduck re-name each biom file [feature_w_tax.biom] to Study_X_feature_w_tax.biom, download and place into our Dropbox directory.